Understanding Computer Vision

2018.02.28

Mr. Calvenn Tsuu

What is the most important sense

- Hearing
- Smell
- Touch
- Taste
- Sight

What is the most important sense

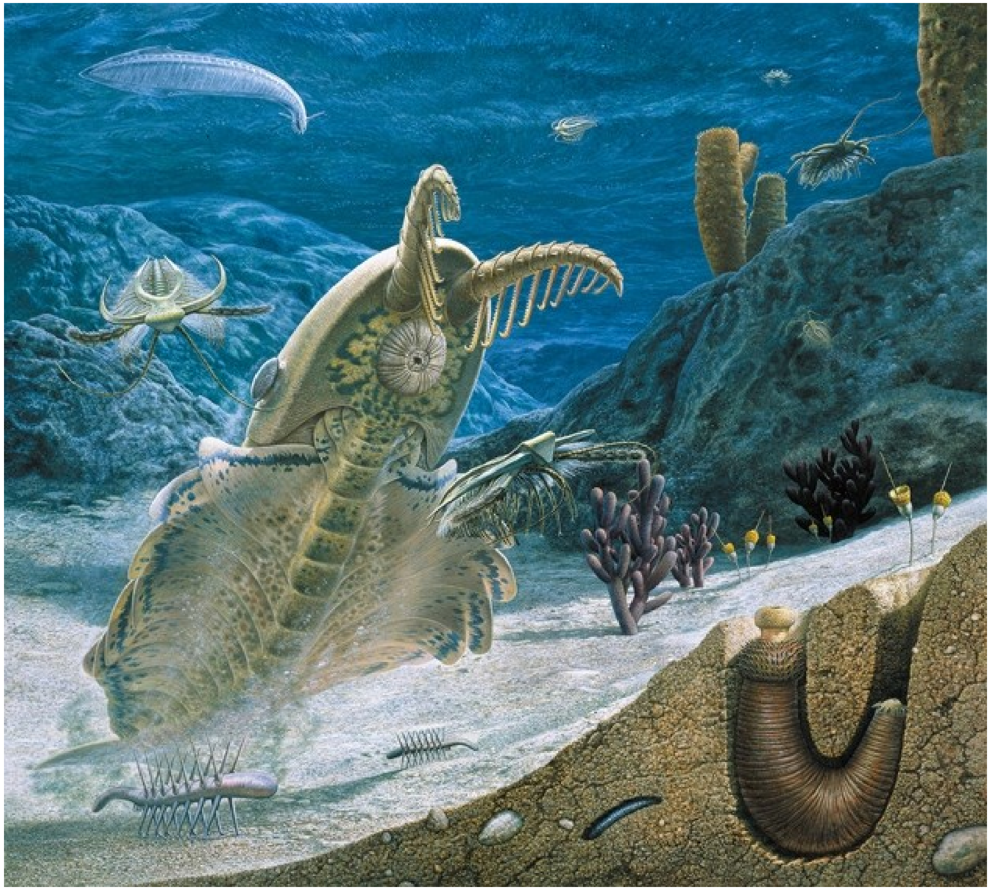
We perceive up to 80% of all impressions by means of our sight.

Processing Speed

Visual information can be processed 60,000 times faster then text, and it is easier to remeber.

Cambrain Explosion

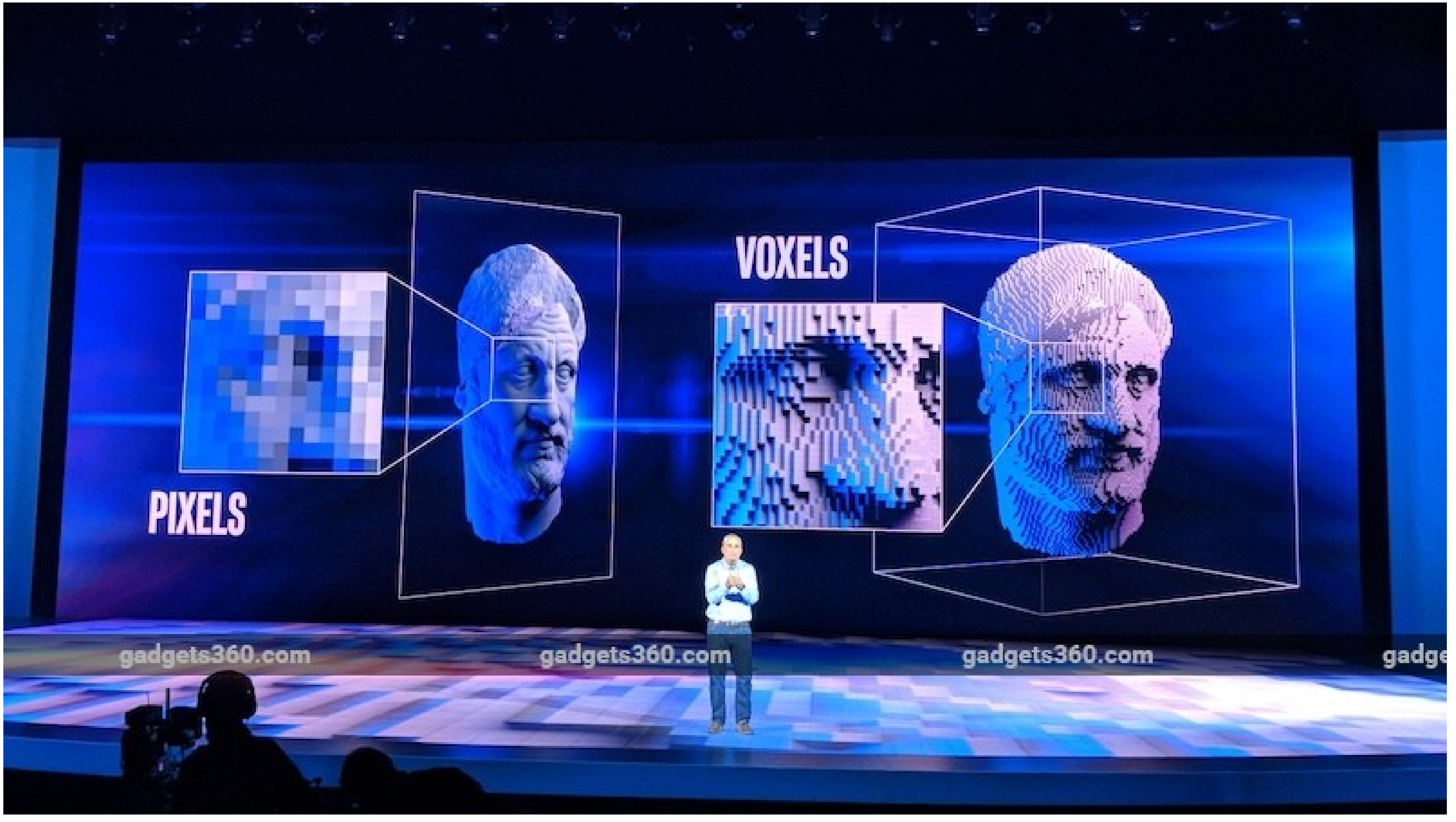
- 500 Million years ago
- Vision is one of the key factor
- Find more food
- Keep them alive
- In short few million years period, total number of animals explosed.



Computer Era

- 16 years ago
- Camera phone captures 640x480 pixels image
- 0.3 megepixels to 15 megepixels
- Tiny screen to Retina screen

Electronic Device Explosion

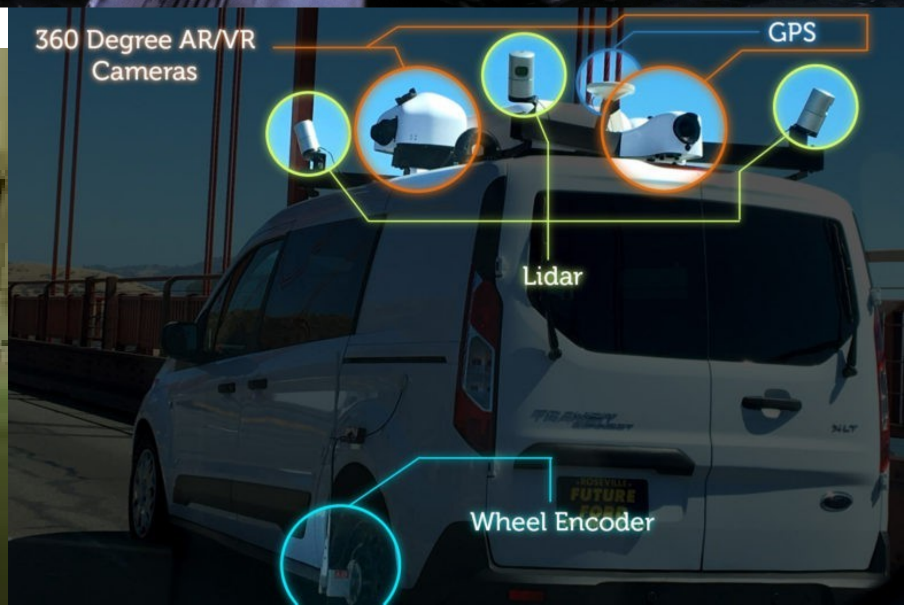
- CES 2018 Las Vegas
- 

Photographer's eye vs Camera

By David Chancellor, National Geographic

Computer Vision

- Surveillance Camera
- Inspection Camera
- Self-driving Car
- Dash Camera
- Speed Camera



“A Critical Moment”

- A moment of decision
- Tipping Point

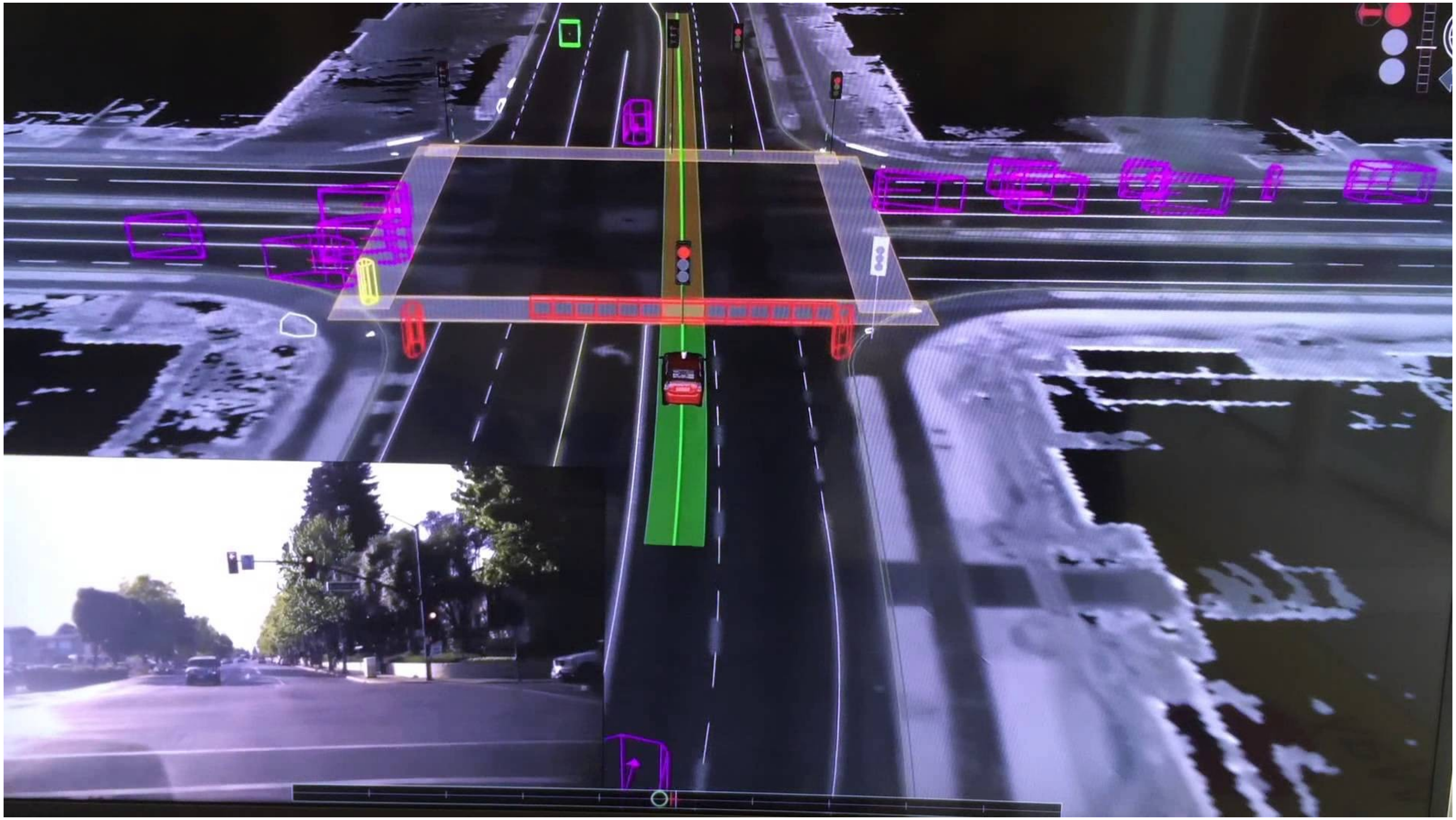
Photographer vs Computer Vision

- Real photographer
- Think before photo taking
- Predict the action
- Critical Moment
- Different Angle and Position
-
- Computer Vision
- Continuious shooting
- Monitoring
- Fixed position

Computer Vision Monitoring

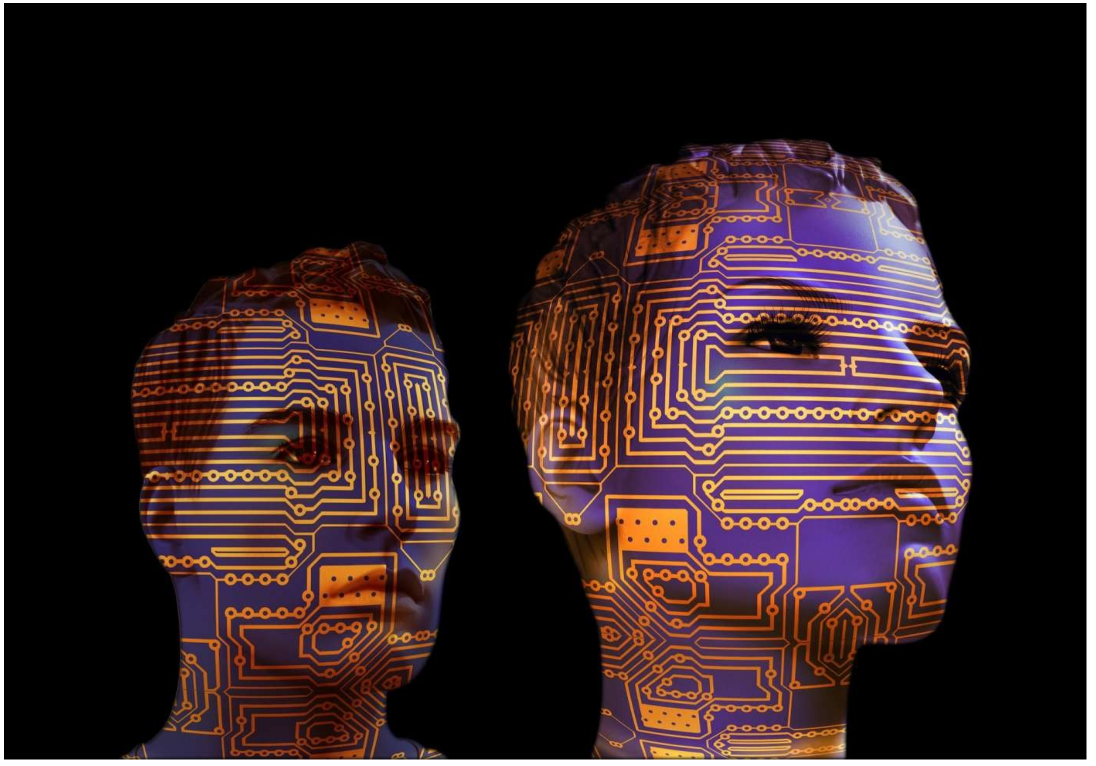
"Find footage of interest using motion based search identity activity patterns with heat mapping."

-Shaw Ads.



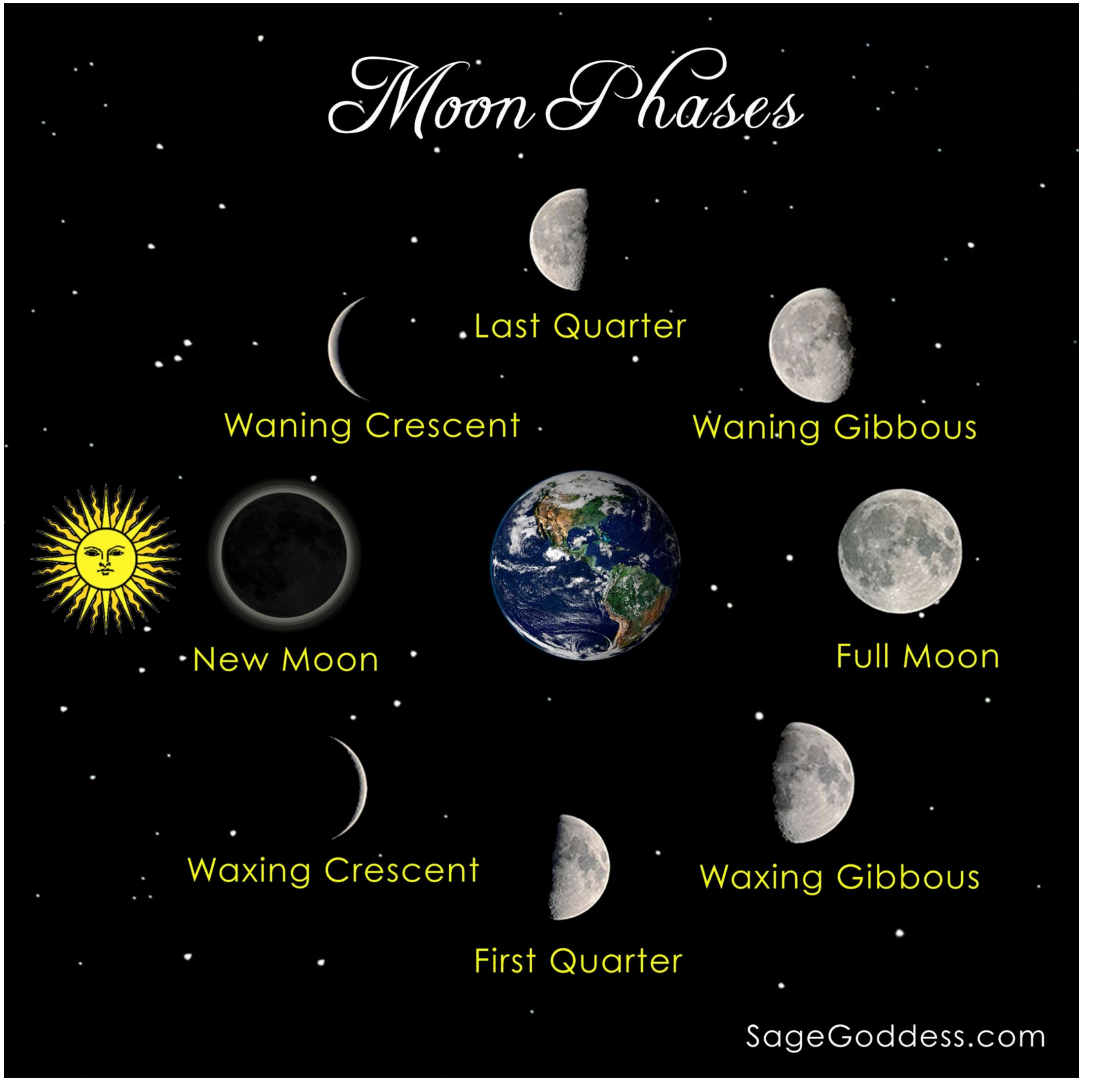
Understanding Computer Vision

- The lens – How the image formed
- The sensor – How the image converted
- The raw data – How the image stored
- Aperture
- Shutter
- Post processing
- - What you can
- get from the
- raw data



Vision is based on Lighting

- Direction
- Intensity
- Reflection
- Transparence

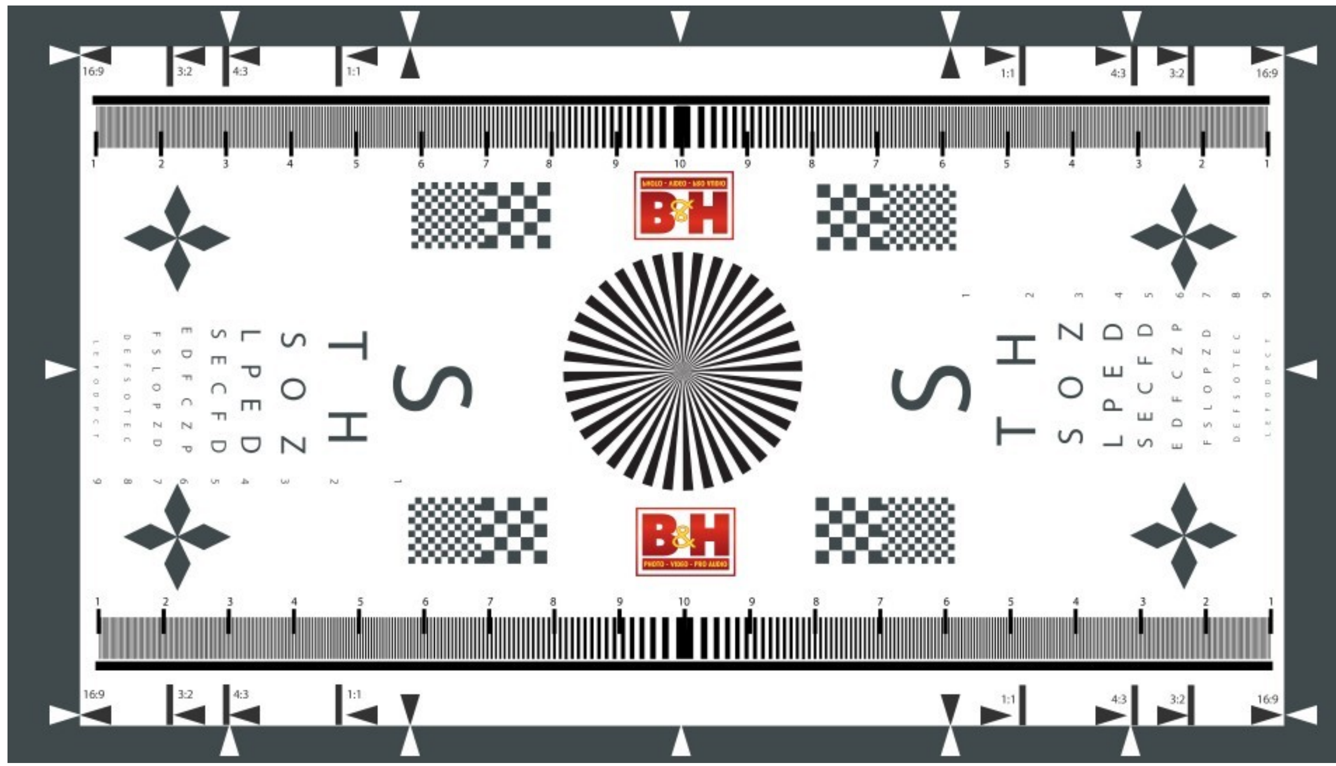


Lighting Intensity/Brightness

- 0.0001 Lux   Moonless, overcast night sky
- 0.05-0.3 Lux Full Moon on a clear night
- 20-50 Lux   Street Light
- 100 Lux   Very dark overcast day
- 250-500 Lux Office work
- 1000 Lux   Drawing/detail work
- 10000 Lux   Not direct sunlight
- 100000 Lux Direct sun
- “Human can easily see from star light to sun light range translate to 30 stops.
- Digital camera is about 2^10 = 1024:1. ”
-
-

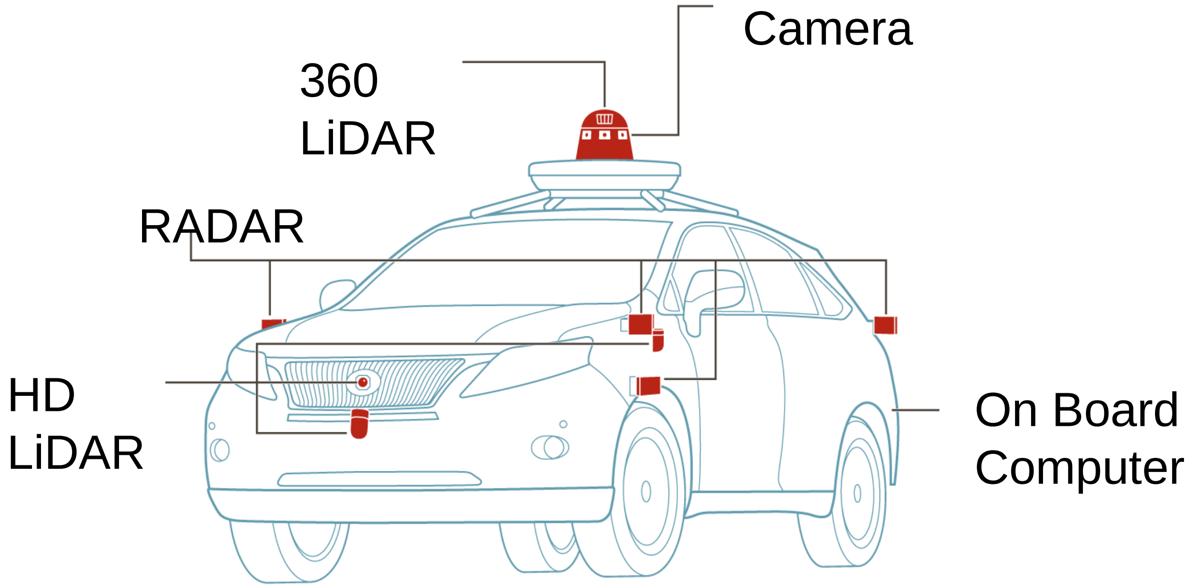
Common Camera Problems

- Near fish eye lens
  - Distortion
- Storage
  - File size
- Dynamic
- Range
- 8~10 stops
- Processing
- Power
- Lens Resolution
  - HD 1920 x 1080 pixels
  -



Why LiDAR is popular

- LiDAR is based on a laser emitting light pluses.
- The brightness level is higher than direct sunlight.
- It always get the front lighting direction.
- Direct Precise Distance Measurement.
- Less effected by the cloud condictions.
- Less post processing required.



LiDAR Challenges

- Low Resolution
- Outdoor performance
  - Dynamic range
  - Weather condition
- Expensive
- Light pollution
- Not possible
- in battle field

Camera based computer vision

- Lens choose:
  - Telescope
  - Fish eye
  - Normal
- Sensor size
  - SD 640x480
  - HD 1920x1080
  - 4K 3840x2160
- Special Cameras
  - Light field camera
  -

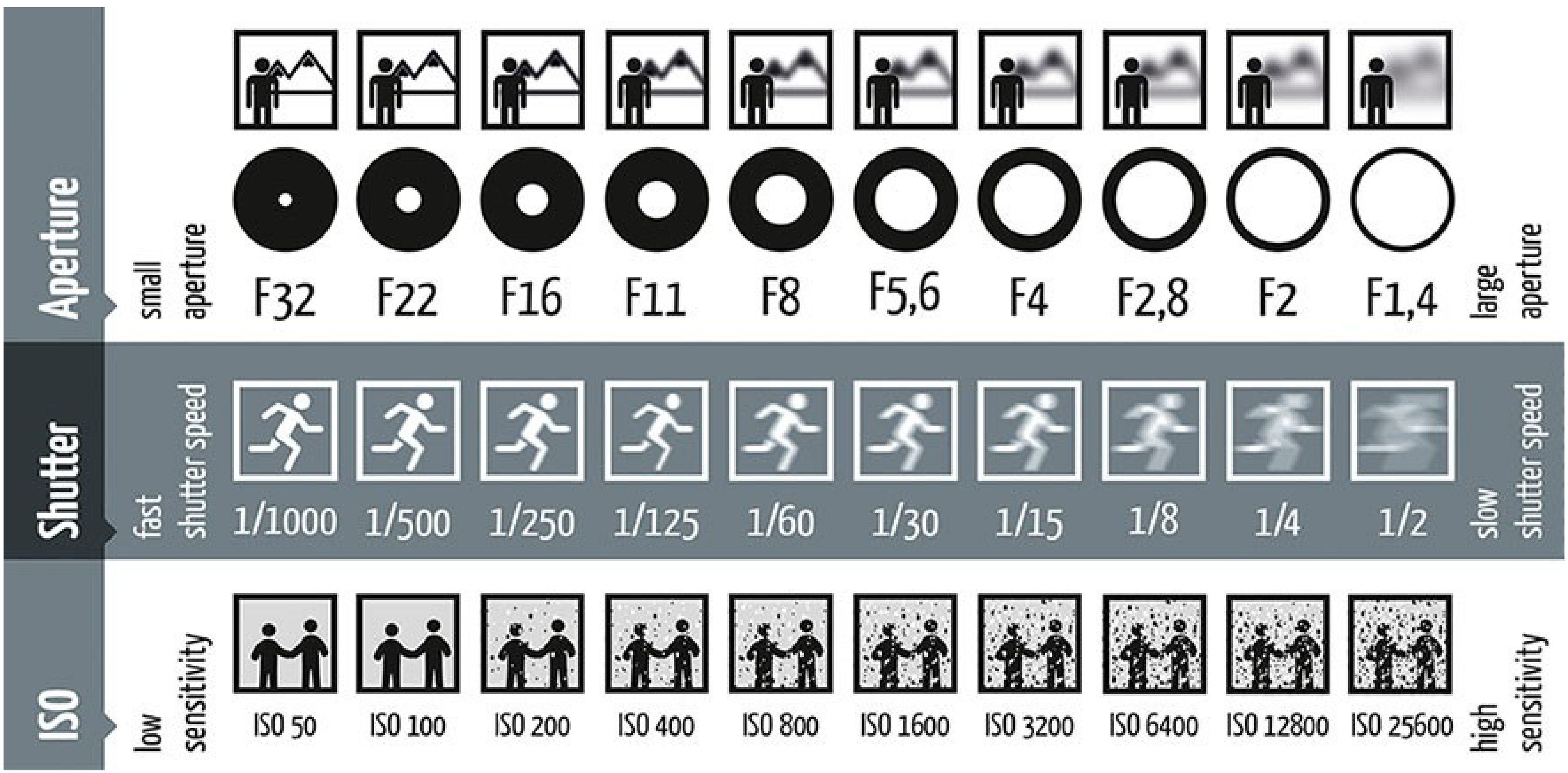
Breakthrough Technologies

- Computing Power – CPU, RAM, GPU
- Deep Learning
  - Neural Networks
  - Tensorflow, etc
- Sensors Fusion
  - LiDAR, RADAR,
  - Camera
- Libraries
  - Computer vision
  - Localization
  - Planning

Camera based CV Challenges

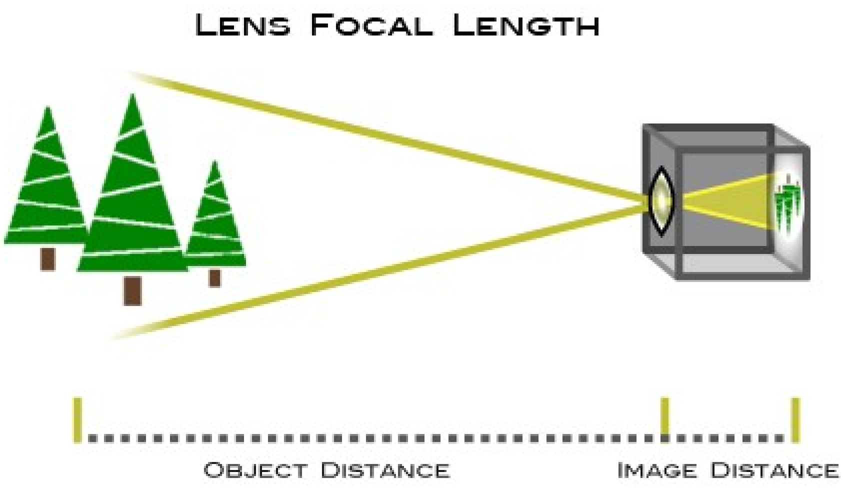
- Very hard to have general solution
- Outdoor performance
  - Dynamic range
- Economic solution for indoor and industry
  - CCTV
  - Inspection
  - Quality Control

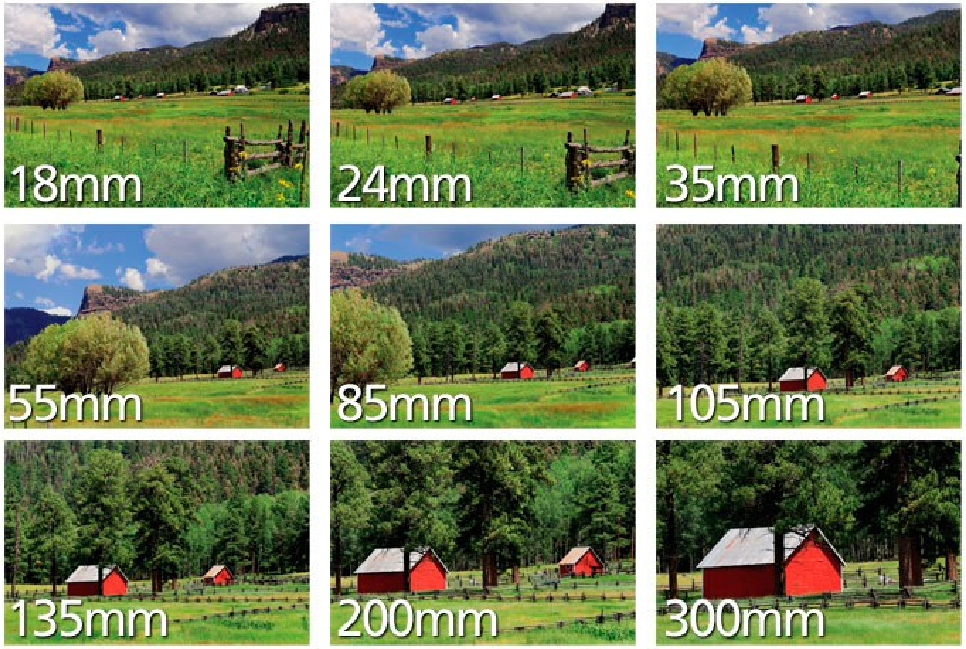
Lens: Shutter and Aperture

- Control the amount of light reaching the sensor
- More light result brighter image reading
- 

Lens: Focal Length and Auto Focus

- Longer focal length, narrower angle of view
- Shorter focal length, wider angle of view
- 35mm -55mm close to human angle of view
- Assume all lens has auto focus, can always get clear picture





https://www.nikonusa.com/en/learn-and-explore/

Lens: Distortion

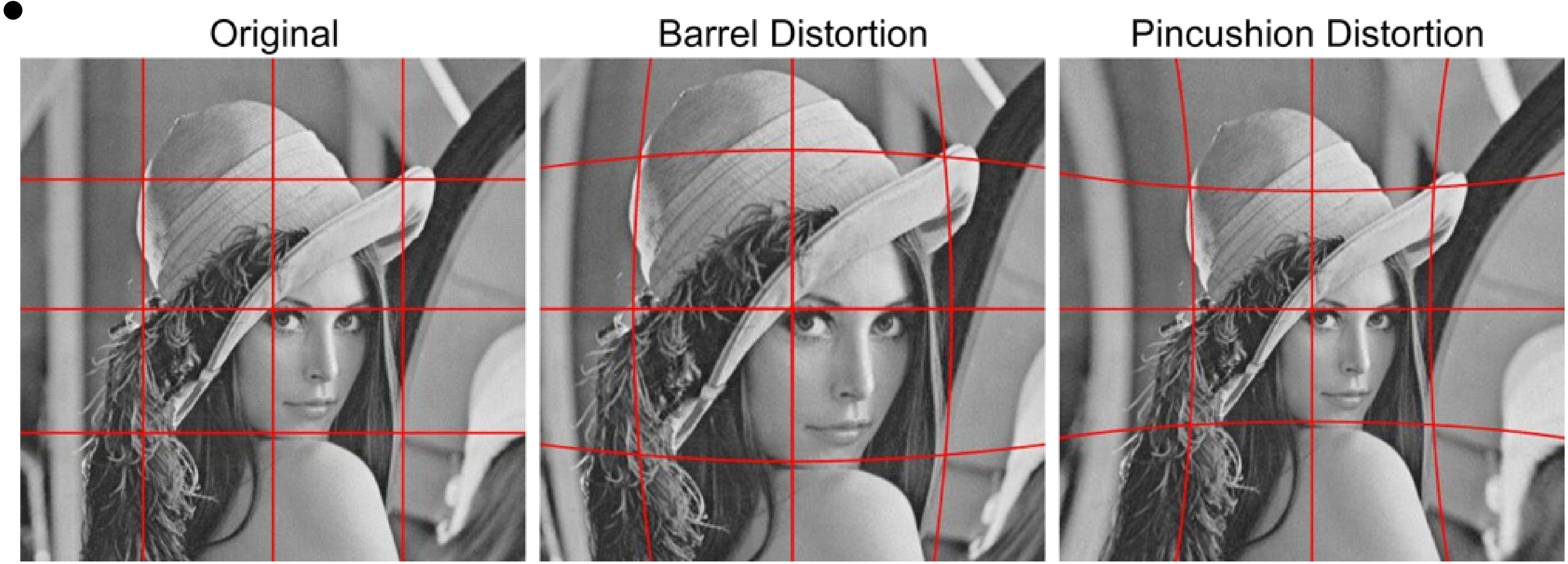
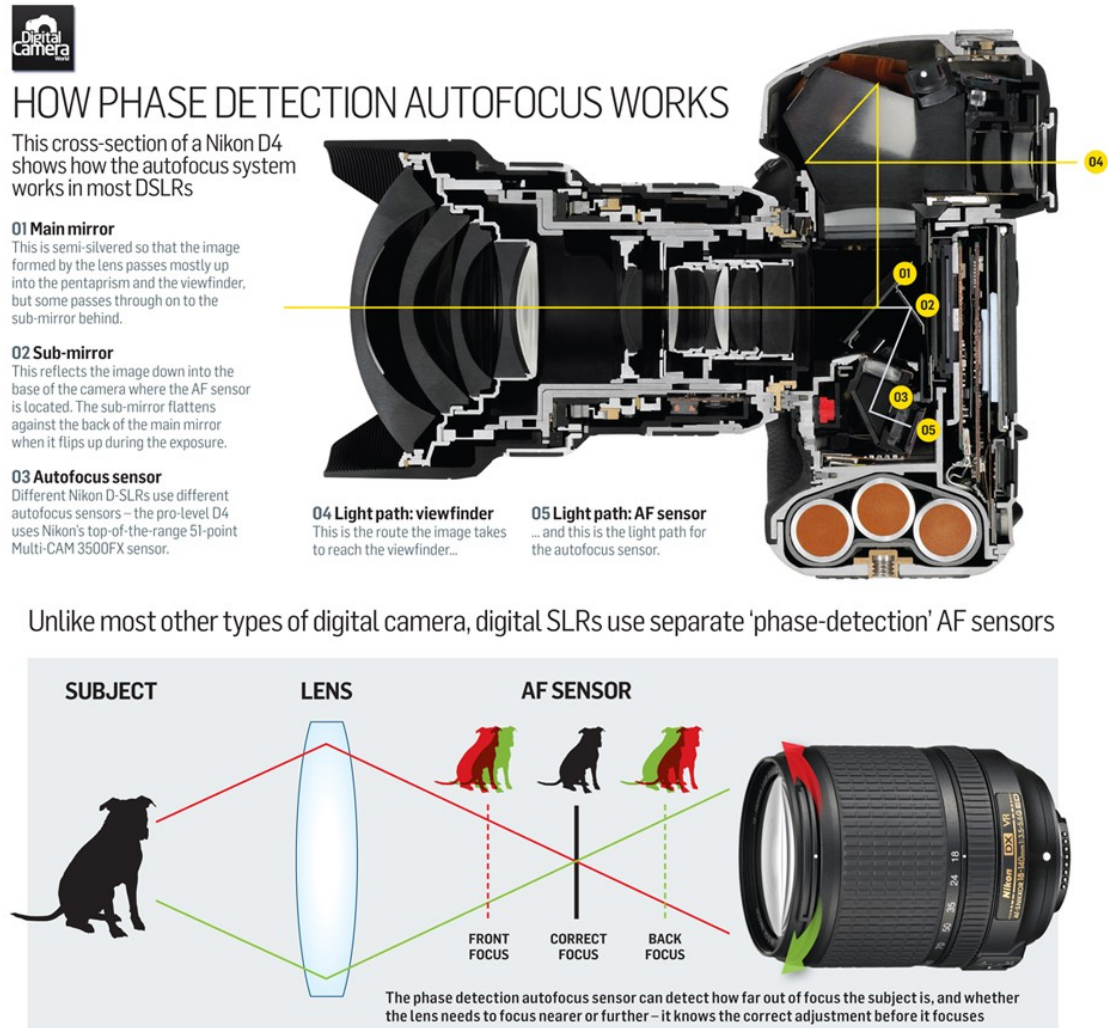
- Barrel distortion
  - Occurrence in wide-angle lenses
- Pincushion distortion
  - Found in low-end telephoto lenses
- 

Image formation: Camera Body

- Don’t need focus
  - Pinhole camera
  - Wide Angle
- Focus before shot
  - Mirrorless camera
  - DSLR: Single
  - Lens Reflection
- Focus after shot
  - Light Field camera

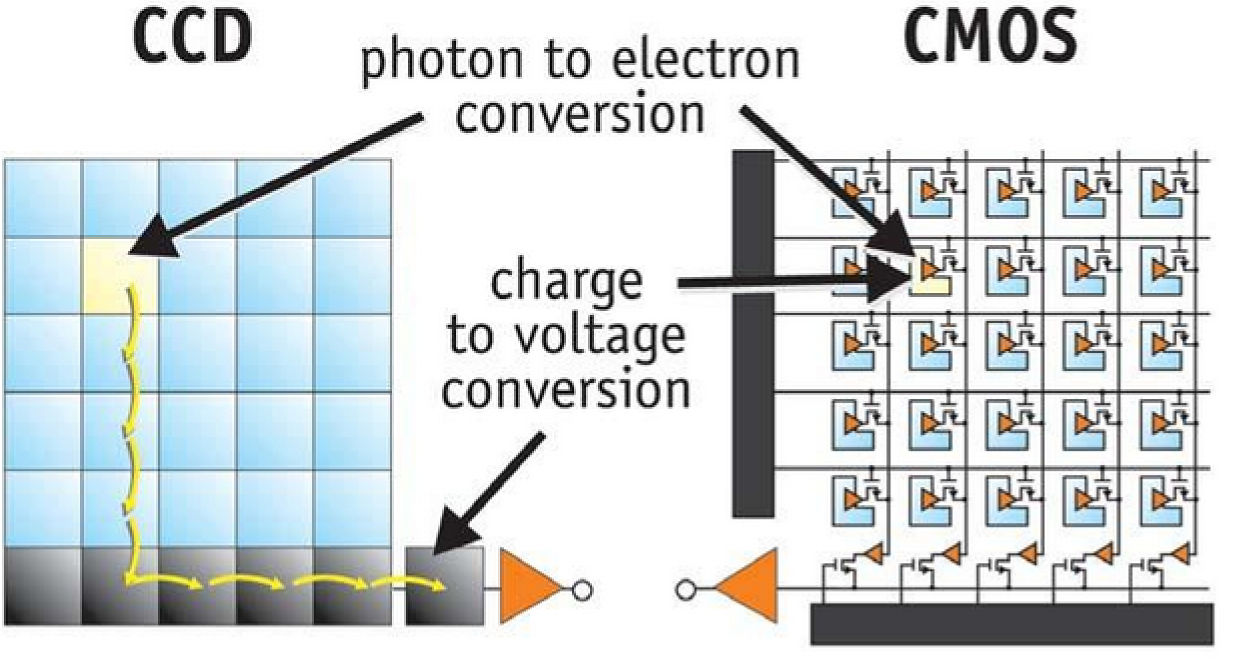


Light Field Camera

- Record at least 4 times more information for each shot
-
- Fast, no auto focusing required before the shot
-
- Low lighting requirement
-
- Depth of field and focus is producted by post processing
-
- 3D images
  -
-

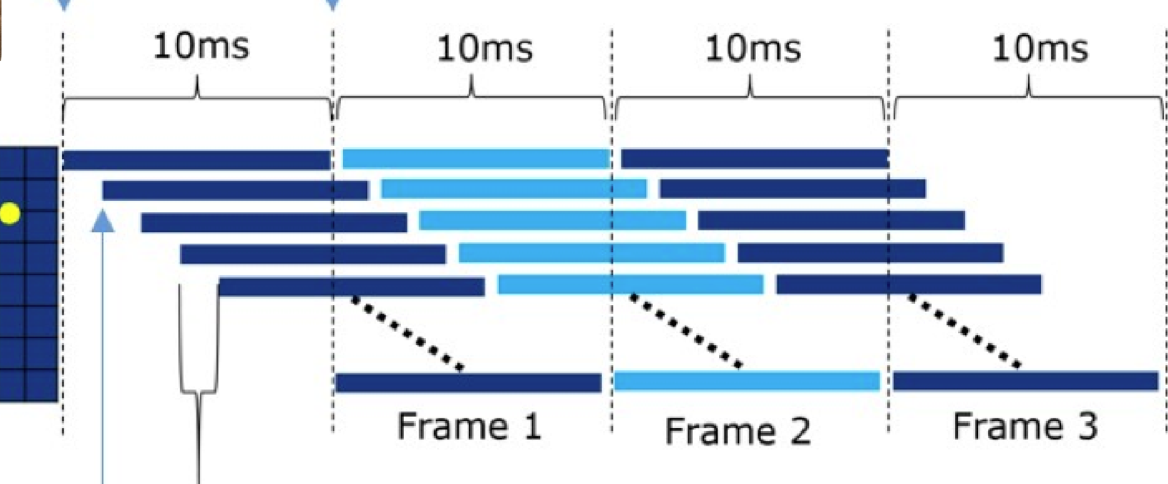
Camera Sensor

- CCD: Charge coupled device
  - Lower noise, expensive, global shutter
  - Higher power consumtion
- CMOS: Complementary Metal Oxide on Silicon
  - Higher noise, cheaper, rolling shutter
  - Very Low power consumtion
  -
-



Global shutter vs Rolling shutter

-

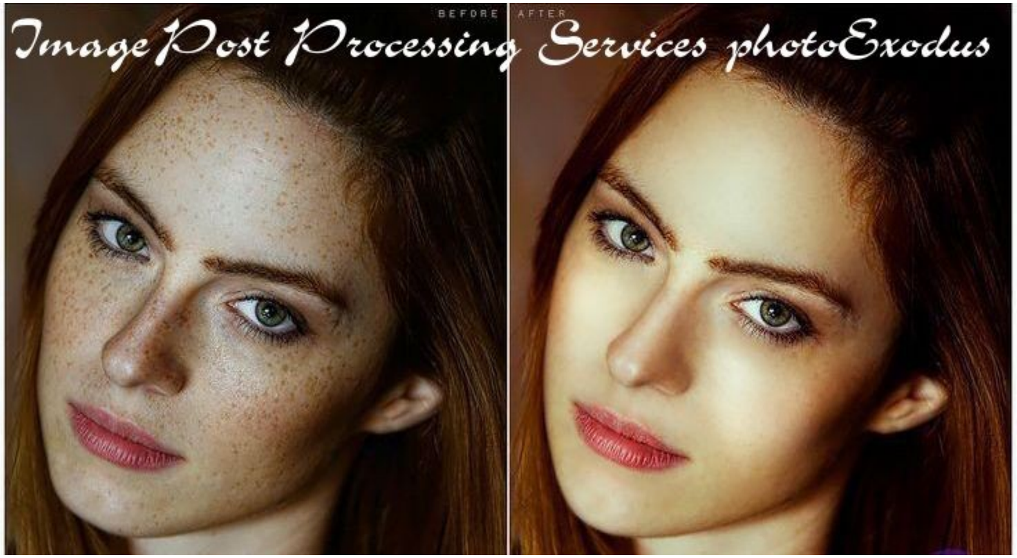


Choose Camera Sensor

- Geometic reconstruction task:
  - The whole image is recorded at exactly the same time
  - Global shutter
- Film industry:
  - Global shutter
- Consumer Market
  - CMOS,
  - Rolling shutter
  - Lower power
  - consumtion
  - Lower cost
-

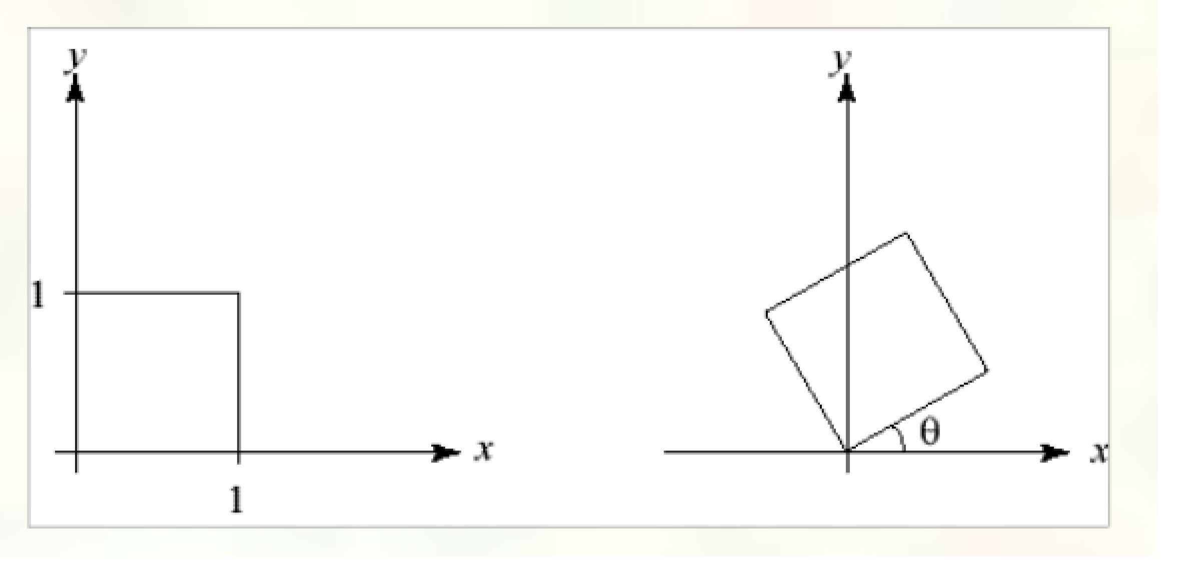
Digital Image Post Processing

- Enhancement
  - Redeye removal
  - Brightness
  - Sharpness
  - Dynamic Range
- Compress file size
- Filtering
- Editing
  - Corp
  - Resize



Transforms in 2D Computer Vision

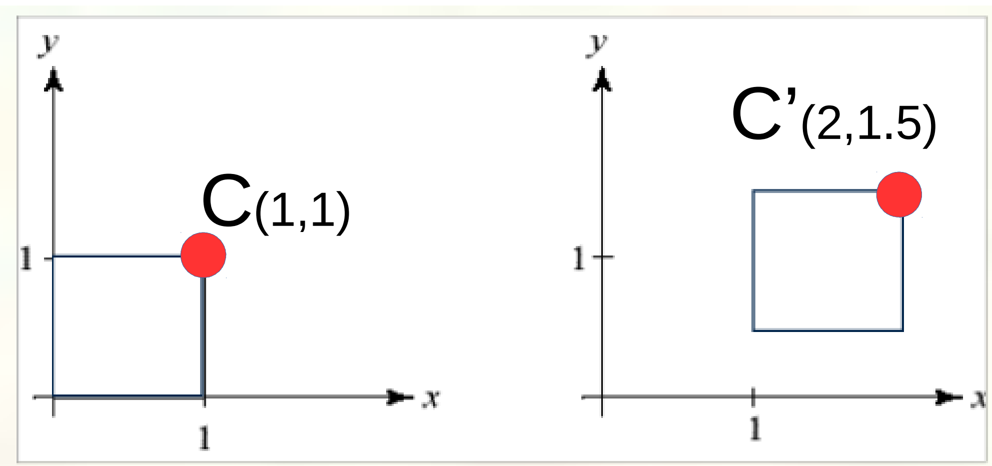
- Affine transforms
  - Translation
  - Rotation
  - Scaling
  - Reflection
  - Shear



-
- Transform between Coordinate systems
  - Local coordinate
  - Global coordinate

2D Transforms: Translation

- Move a point a = 1 in x direction
- Move a point b = 1/2 in y direction
- Translation Matrix T(a,b) =

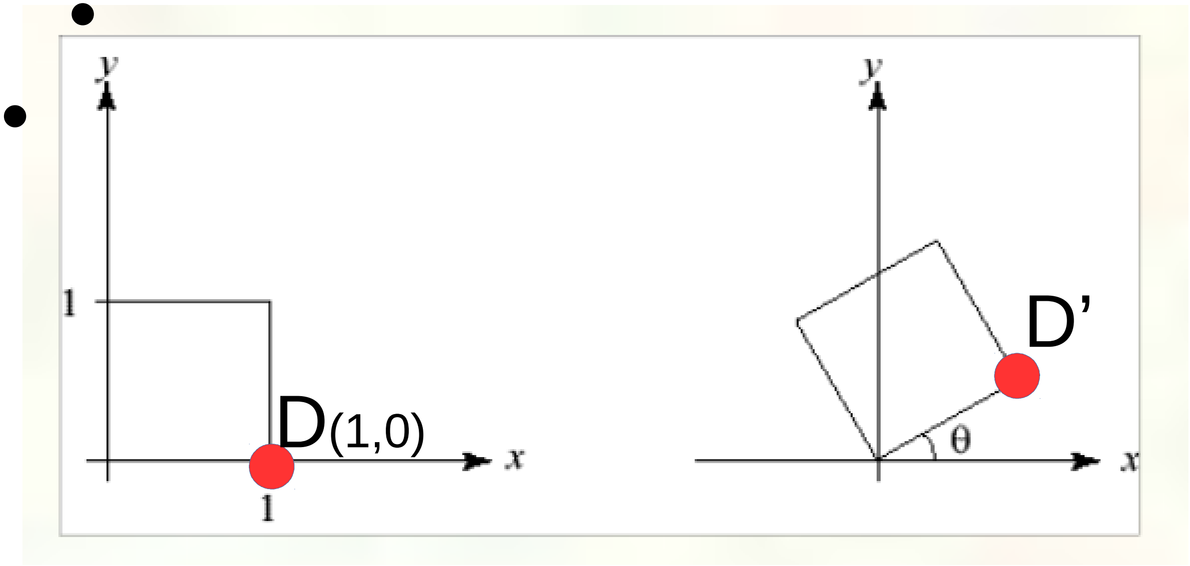


$$\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1/2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

- Move a point C(1,1) to C'(2, 1.5)
  - Matrix multiplication
- Loop through all points if needed

2D Transforms: Rotation

- Based on Origin (0,0),
- Counter-clock wise by angle theta
- Rotation Matrix R(theta) =
  -
-

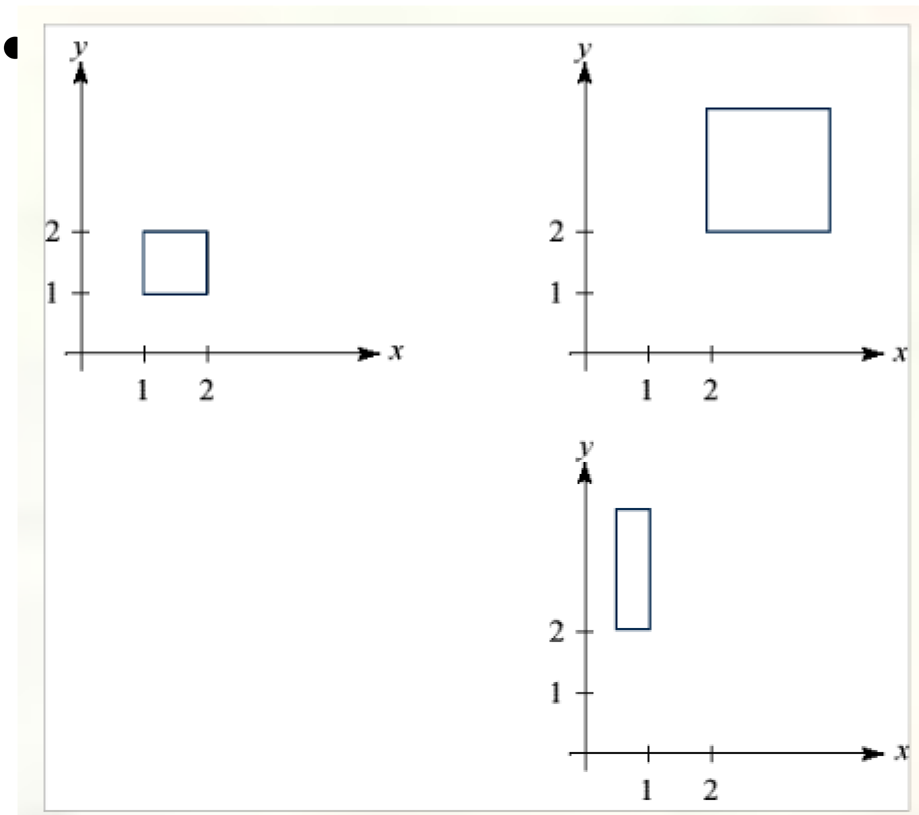


$$\begin{bmatrix} \cos(\theta), & -\sin(\theta), & 0 \\ \sin(\theta), & \cos(\theta), & 0 \\ 0, & 0, & 1 \end{bmatrix}$$

  - Move a point D(1,0) to D'
    - Matrix multiplication
  - Loop through all points if needed
    -

2D Transforms: Scaling

- Scaling Matrix S(s,t) = $\begin{bmatrix} s, & 0, & 0 \\ 0, & t, & 0 \\ 0, & 0, & 1 \end{bmatrix}$
  -



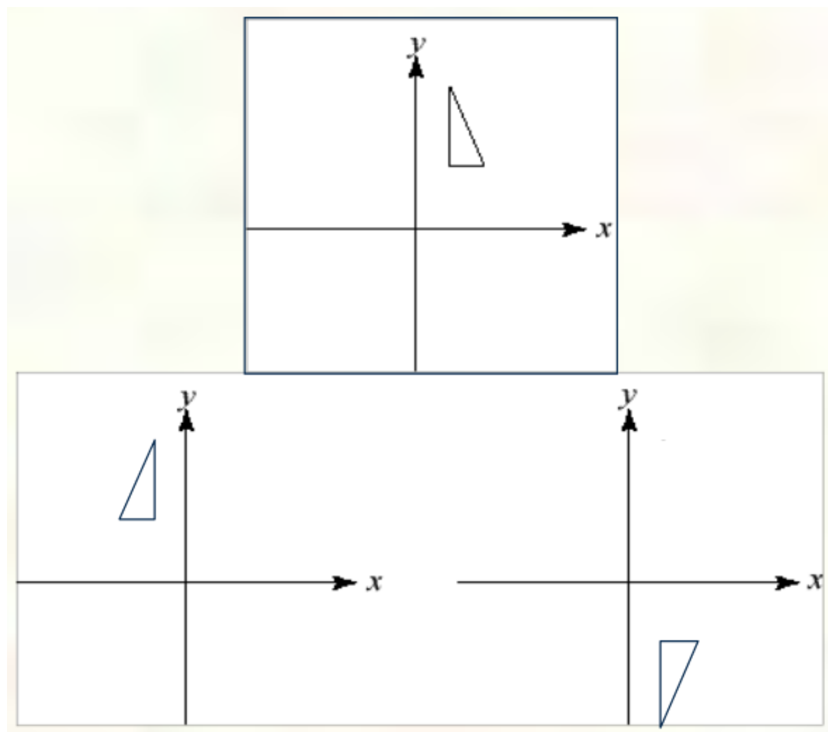
$\begin{bmatrix} 2, & 0, & 0 \\ 0, & 2, & 0 \\ 0, & 0, & 1 \end{bmatrix}$

$\begin{bmatrix} 0.5, & 0, & 0 \\ 0, & 2, & 0 \\ 0, & 0, & 1 \end{bmatrix}$

  -
  - Loop through all points if needed
    -

2D Transforms: Reflection

- Refleted about x or y axis Matrix



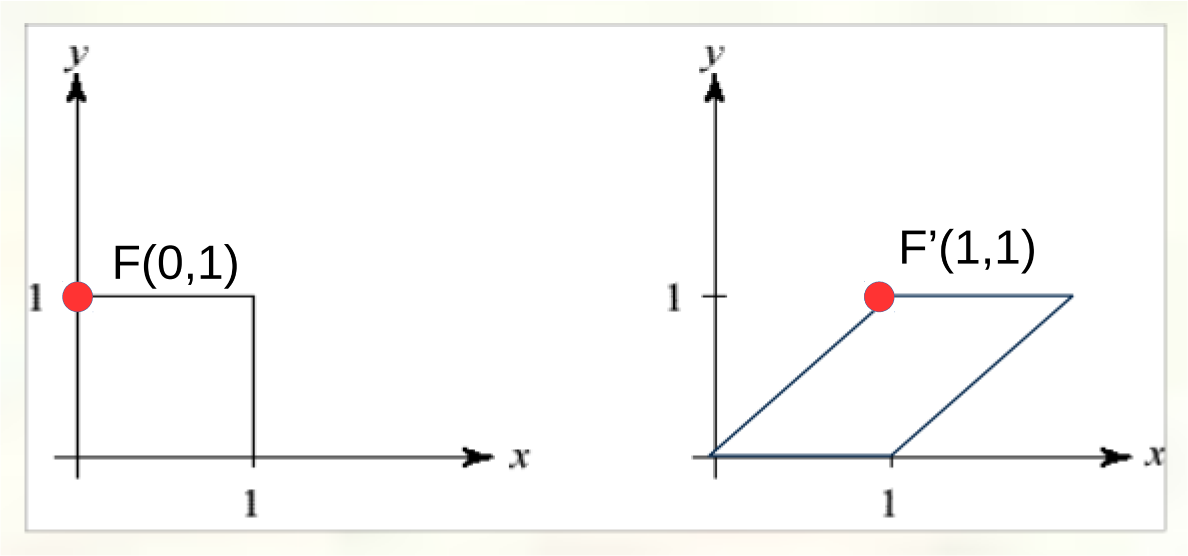
$$M(y) = \begin{bmatrix} -1 & 0, & 0 \\ 0, & 1, & 0 \\ 0, & 0, & 1 \end{bmatrix}$$

$$M(x) = \begin{bmatrix} 1, & 0, & 0 \\ 0, & -1, & 0 \\ 0, & 0, & 1 \end{bmatrix}$$

- Loop through all points if needed

2D Transforms: Shear

- Use a to control x axis shear
-



$$S(x) = \begin{bmatrix} 1, & a, & 0 \\ 0, & 1, & 0 \\ 0, & 0, & 1 \end{bmatrix}$$

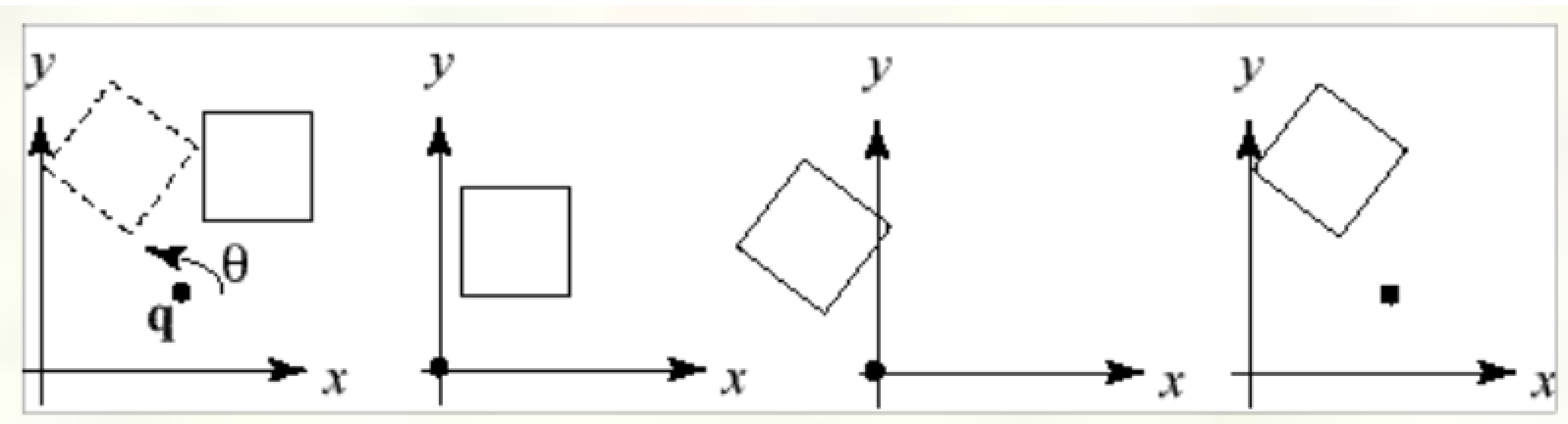
a = 1

$$S(y) = \begin{bmatrix} 1 & 0, & 0 \\ b, & 1, & 0 \\ 0, & 0, & 1 \end{bmatrix}$$

- Use b to control y axis shear
- Loop through all points if needed
  -
-

Composition of Transforms

- Rotation about arbitrary points
- Given rotation point q = [qx, qy, 1]T
-



-
  - Translate q to origin
  - Rotate
  - Translate back
  - T(q)R(theta)T(-q)q
  -
  - Different multiplication order lead to different results
    -
-

Local/Global Coordinates

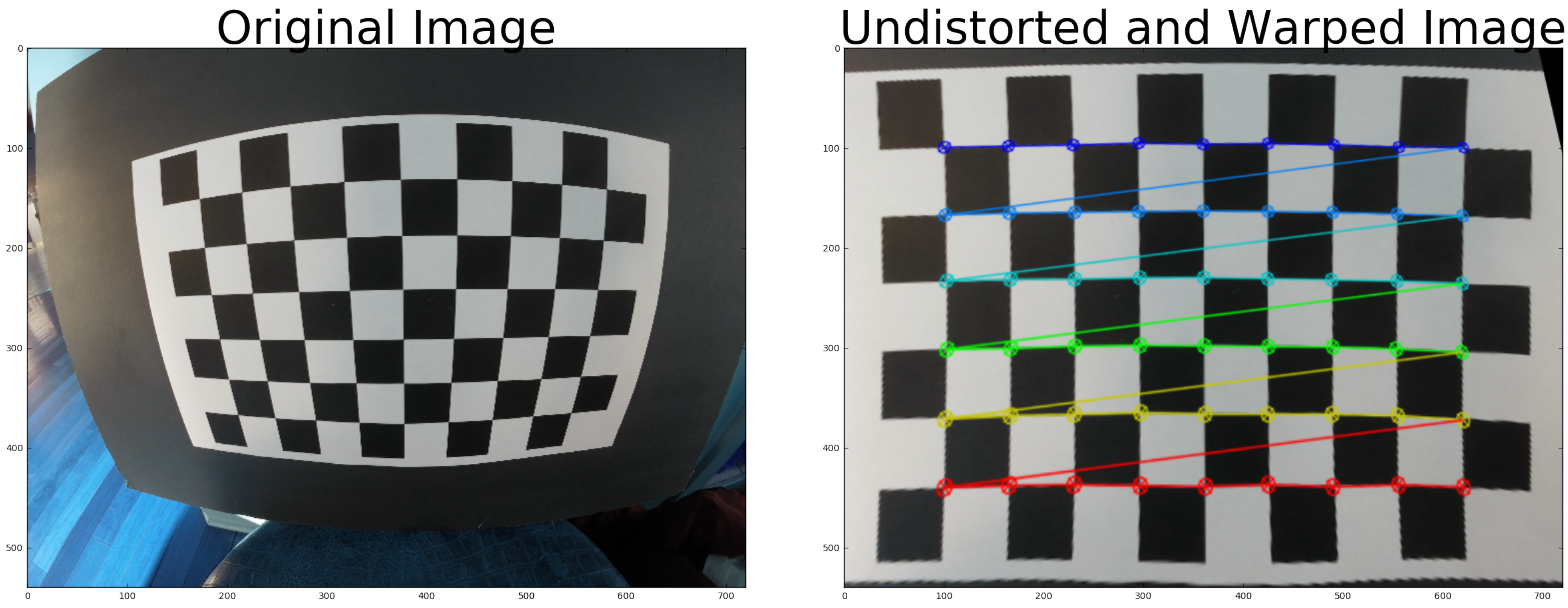
- Global coordinate: GPS waypoints
  - Longitude: -114.185893
  - Latitude: 51.035545
- Local coordinate: Car Camera, LiDAR
  - Car always at origin(0,0)
  - All measure refer to car origin
  - Given car global location(x, y) and
  - heading direction (theta)
  - Convert GPS waypoints into local coordinate
- Now you can do all the magic!!!

Practical Computer Vision

- Barrel Distortion Issues:
- Closer object appears larger than it should be
- Farther object appears even farther that it should be
- If you need use image to estimate distance and geometary, you need consider the distortion.

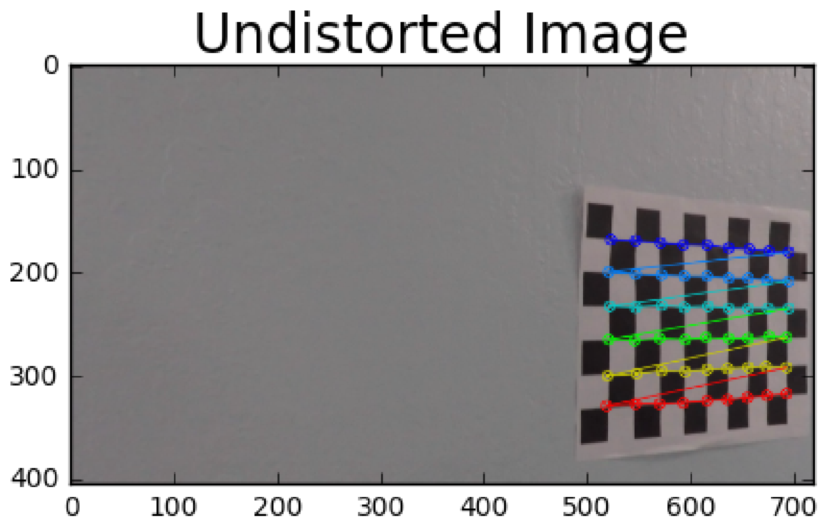
Camera Calibration: OpenCV tools

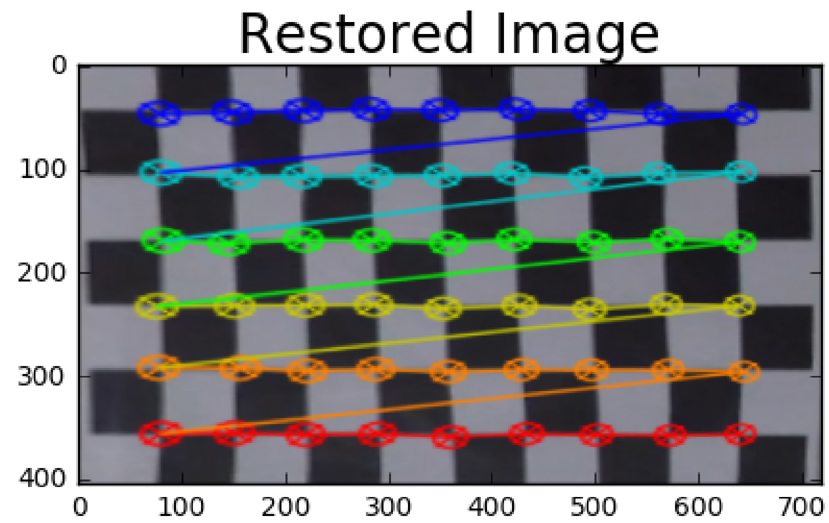
- cv2.findChessboardCorners()
- cv2.calibrateCamera()
- cv2.undistort()
- cv2.warpPerspective()



Camera Calibration: OpenCV Tips

- Take 20-30 pictures for calibration to cover all image areas and angles
- Each resolution has its own matrix
- Save the matrix for future use



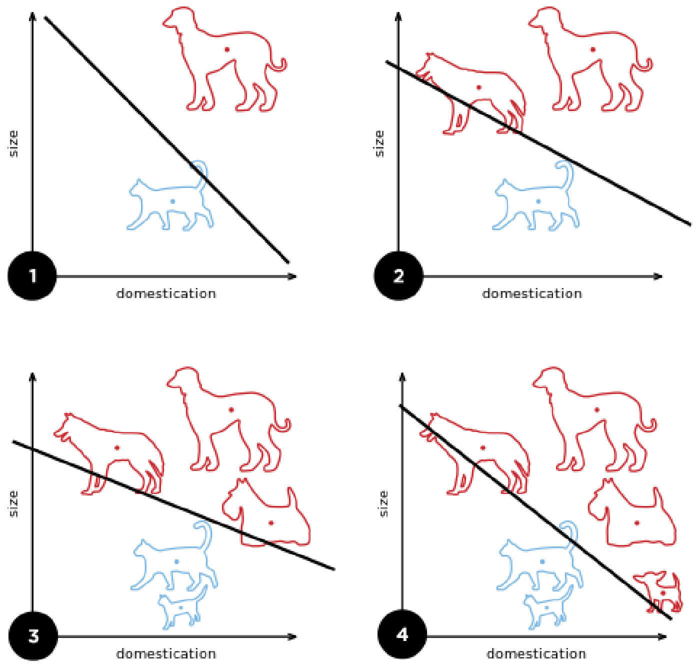


Traditional CV vs Deep Learning CV

- How to tune the filter parameters
- Speed
- Training
- Power

Early Artifical Neural Networks

- 1986 Geffrey Hinton find the way to train the multi-layer neural network
  - Backpropagation(Backward propagation of errors)



Brutal Force

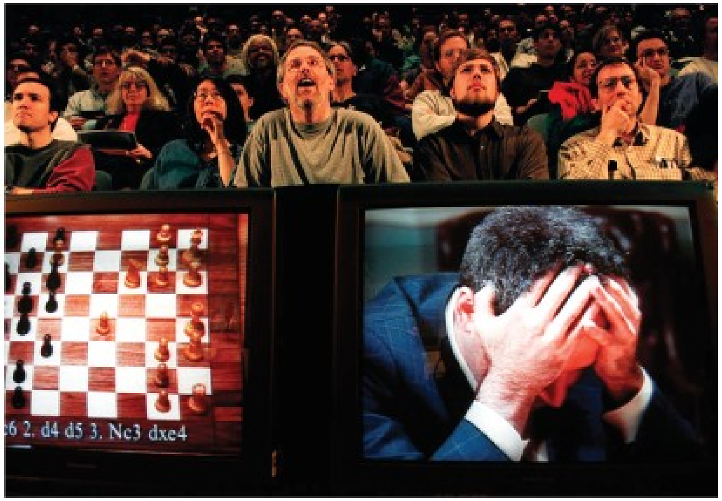
KEY MOMENTS IN DEEP-LEARNING HISTORY 1989-1997

1989
French researcher Yann LeCun, then at Bell Labs, begins foundational work on a type of neural net that becomes crucial for image recognition.

1991
German researchers Sepp Hochreiter and Jürgen Schmid-huber pioneer a neural net with memory features, which eventually proves superior for natural-language processing.

1997
IBM's Deep Blue beats **world champion Garry Kasparov** (right) in chess using traditional AI techniques.

→



STAN HONDA—AFP/GETTY IMAGES

Dataset

KEY MOMENTS IN DEEP-LEARNING HISTORY 1990's-2011

Mid-1990s
Neural nets fall into disfavor again, eclipsed by other machine-learning techniques.

2007
Fei-Fei Li founds ImageNet and begins assembling a database of 14 million labeled images that can be used for machine-learning research. →

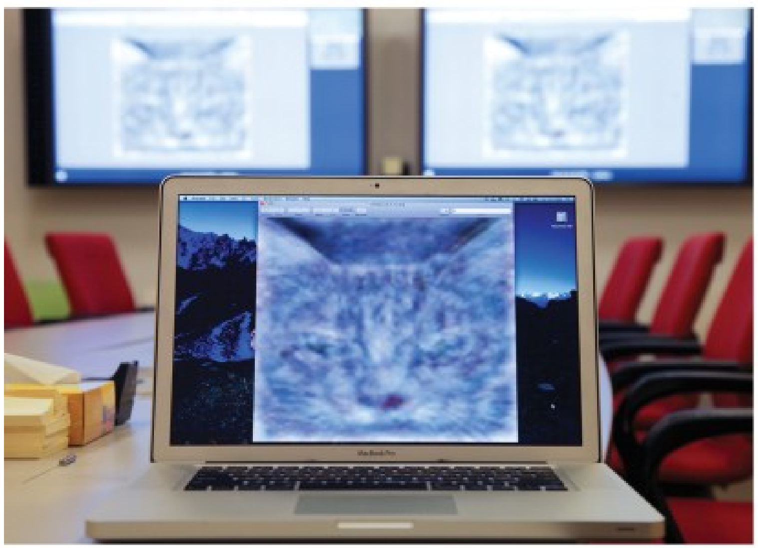
2011
Microsoft introduces neural nets into its speech-recognition features.

2011
IBM's Watson beats two champions at Jeopardy using traditional AI techniques.

CARLOS CHAVARRIA—THE NEW YORK TIMES/REDUX PICTURES

Object recognizing

KEY MOMENTS IN DEEP-LEARNING HISTORY 2012-2013



2012
JUNE
Google Brain publishes the "cat experiment." A neural net, shown 10 million unlabeled YouTube images, has trained itself to recognize cats.

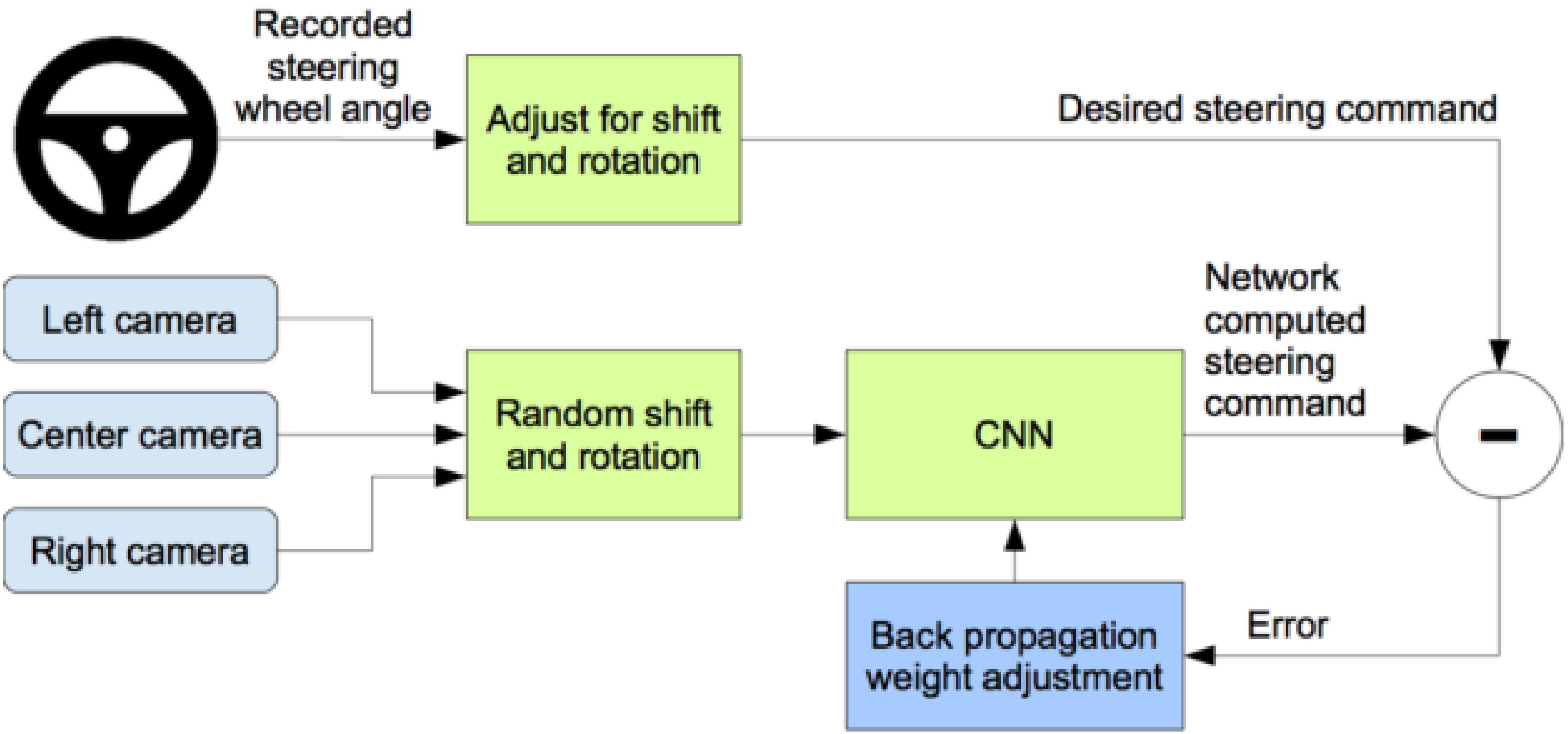
AUGUST
Google introduces neural nets into its speech-recognition features.

OCTOBER
A neural net designed by two of Hinton's students wins the annual ImageNet contest by a wide margin.

2013
MAY
Google improves photo search using neural nets.

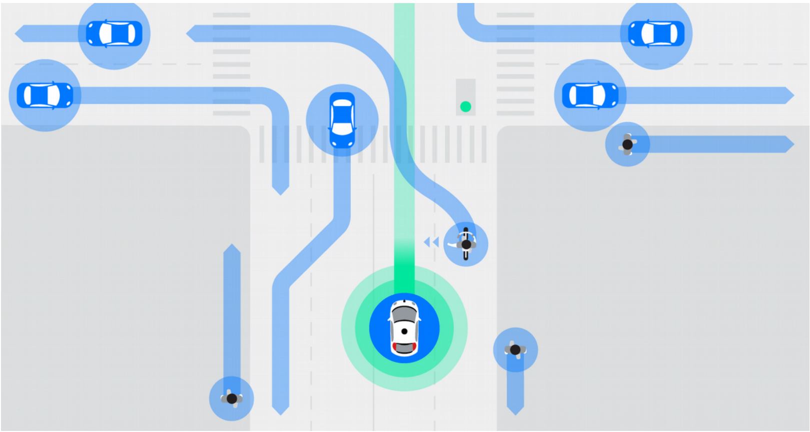
JIM WILSON—THE NEW YORK TIMES/REDUX PICTURES

End to End Learning



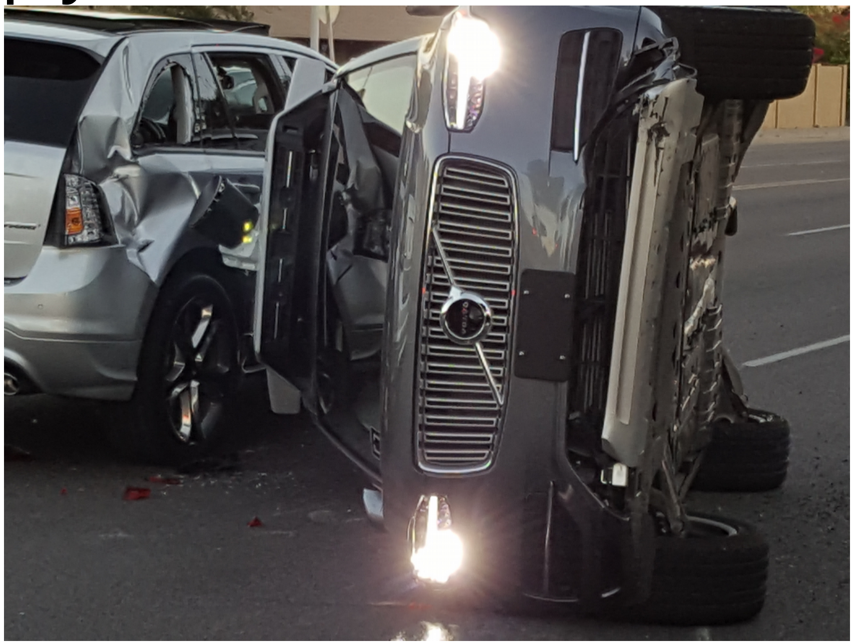
Environment

- In real situation:
- City/Highway/Traffic/Construction/Emergency
- Weather/Rain/Snow/Day/Night

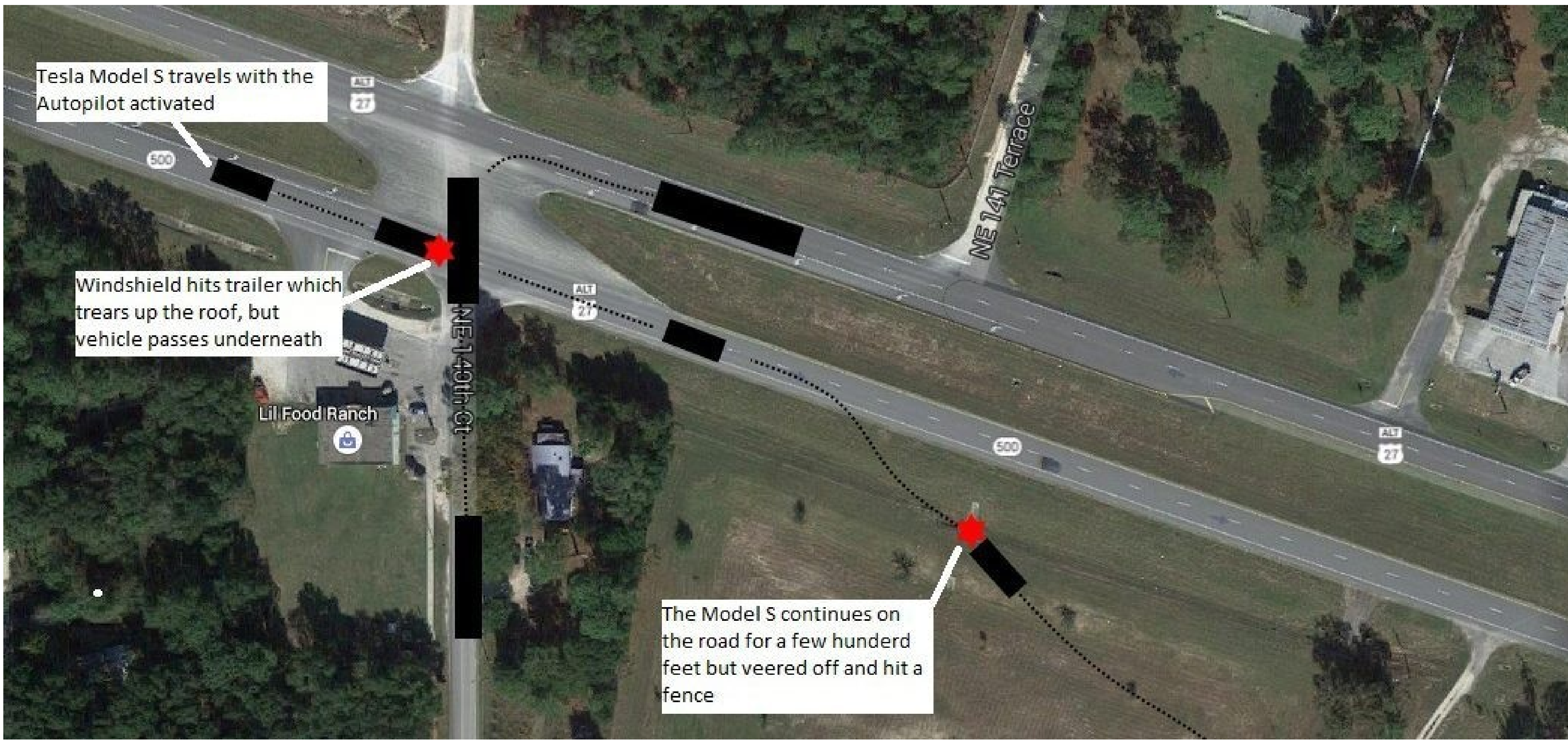
 

Compansate/Fall Safe

- System Failures:
  - Hardware failure 10%
  - Software bugs 70%
  - Unexpected event 20% / not able to generalize the machine learning and apply to new events.
- Plan B
  - Redundancy
  - Fall safe
  - Reboot
- Call for help



Tesla/Semi Fatel Crash



Can you see the truck

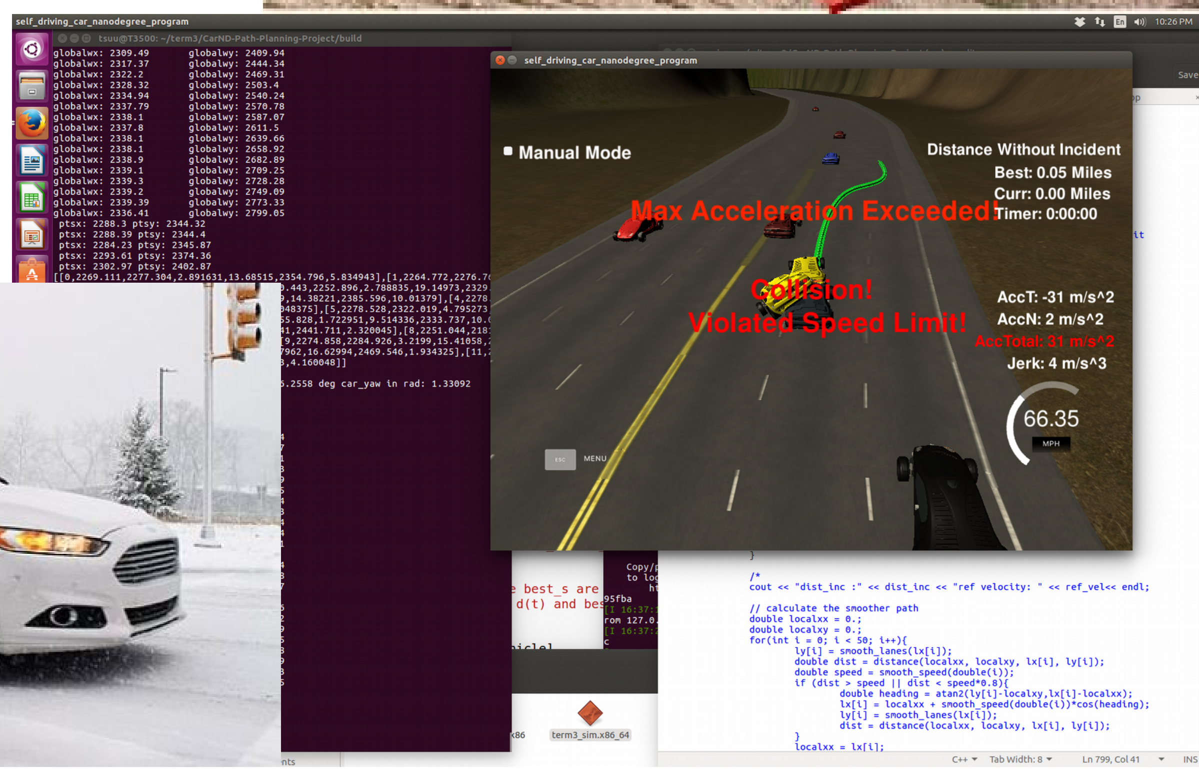
- Tesla can’t while the accident happened

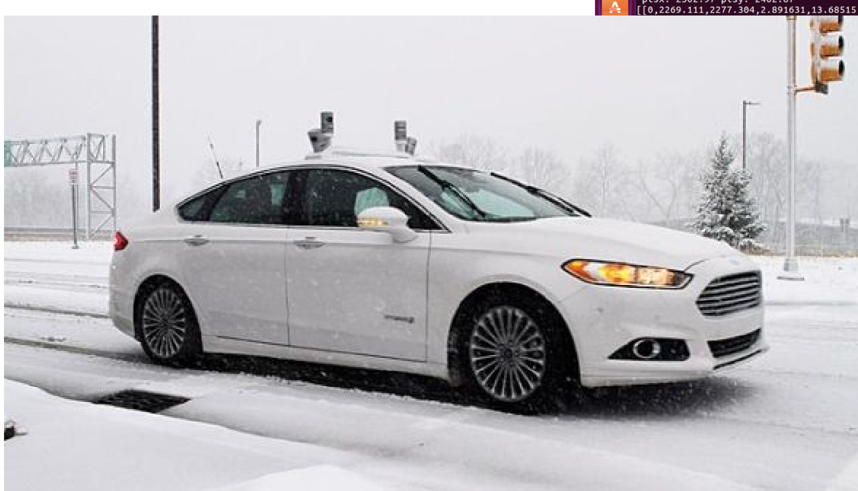
SAE Autonomy Levels

	Name	Definition	Execution	Monitoring	Fallback	System Capability
0	No Automation		Human	Human	Human	N/A
1	Drive Assistance	Hands on	Human /system	Human	Human	ACC, LKA, Parking
2	Partial Automation	Hands off (mandatory check in)	System	Human	Human	Testing stage
3	Conditional Automation	Eyes off	System	System	Human (traffic jam)	Stop and go
4	High Automation	Mind off In optimital areas (geofencing, traffic jam, parking lot etc)	System	System	System (abort the trip if driver does not retake controls	Robot suttle bus
5	Full Automation	Steering Wheel optional	System	System	System	Robot taxi

Technical Challenges

- Computer vision
- Sensor Fusion
- Localization
- Path Planning
- Controller



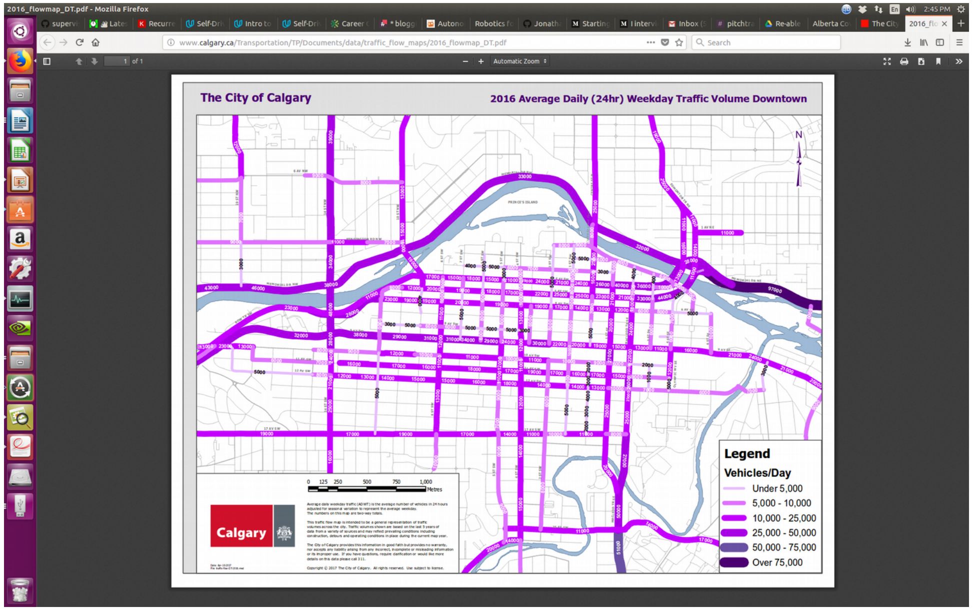


Solutions - Perception

- Traditional Computer vision
- Deep learning object detection and recoganation
  - Traffic signs
  - Traffic lights
  - Objects

Our Focued Area

- Neural Network design and test
- Create training and validation dataset
- Produce control units
- User experience pilot project



Building Self Driving Car - Local Dataset - Night

Current Plan

- Launch Self-Driving User experienment platform prototype 1 and 2 early 2018
- Approach Industrial leader for private testing
- Approach Provincial government and City for public testing permit
- Recent work:
  - https://www.youtube.com/watch?v=7nwrwjLZxcw
-

Team Leader Profile

- Calvenn Tsuu P. Eng PMP
- Self-Driving Car Engineer
- Mechanical Engineer

- Experiences in Robotics, computer vision, Neural Network, CNC machine tools, Automation and Project Management https://www.linkedin.com/in/calvenn-tsuu-94a90317/

- Phone: 403-926-5689
- Email: calvenn.tsuu@gmail.com